Temporal variability in spontaneous Hungarian speech

Beke, András – Gósy, Mária – Horváth, Viktória

Research Institute for Linguistics, Hungarian Academy of Sciences
33 Benczúr Street, Budapest, Hungary
{beke.andras,gosy.maria,horvath.viktoria}@nytud.mta.hu

Abstract

The aim of this paper is an objective presentation of temporal features of spontaneous Hungarian narratives, as well as a characterization of separable portions of spontaneous speech (thematic units and phrases). An attempt is made at capturing objective temporal properties of boundary marking. Spontaneous narratives produced by ten speakers taken from the BEA Hungarian Spontaneous Speech Database were analyzed in terms of hierarchical units of narratives (duration of pauses and of units, speakers' rates of articulation, number of words produced, and the interrelationships of all these). The results confirm the presence of thematic units and of phrases in narratives by objective numerical values. We conclude that (i) the majority of speakers organize their narratives in similar temporal structures, (ii) thematic units can be identified in terms of certain prosodic criteria, (iii) there are statistically valid correlations between factors like the duration of phrases, the word count of phrases, the rate of articulation of phrases, and pausing characteristics, and (iv) these parameters exhibit extensive variability both across and within speakers.

Keywords: articulation tempo, pauses, thematic units, phrases

1. Introduction

Temporal characteristics of spontaneous speech are affected by a number of factors. The aim of the present study is an objective presentation of temporal features of spontaneous narratives including a characterization of the phrases in the narratives. An attempt is made at defining various units of spontaneous narratives and capturing objective acoustic-phonetic properties of boundary marking. We try to identify the factors determining the articulation rate of portions of speech within and across speakers and to find out whether the acoustic-phonetic parameters we analyze make up a characteristic pattern, and if they do, how they can be described.

Klatt (1976) listed seven factors that determine the temporal patterns of speech: extralinguistic factors (the speaker's mental or physical state), discourse factors (position within discourse), semantic factors (emphasis and semantic novelty), syntactic factors (phrase-final lengthening), morphological factors (word-final lengthening), phonological and phonetic factors (stress, phonological length distinctions), and physiological factors (segment-internal temporal structure). Additional factors may also play a role, like topic of discourse, speech type, speech situation, speech partner (Youan et al., 2006). An analysis of tempo in Dutch interviews confirmed the distinct role of phrase length (Quené, 2005). Dialect also seems to be a crucial factor, as shown by an analysis of speech rate in 192 speakers of American English from Wisconsin and North Carolina (Jacewicz et al., 2010). Similar results emerged from an analysis of 267 hours of spontaneous dialogues produced by Dutch speakers living in the Netherlands and in Belgium (Verhoeven et al., 2004). Both of the last-mentioned papers claim, in addition, that men tend to speak faster than women do, and that young speakers' speech rate is faster than that of older speakers. Some data gathered from speakers of (American) English partly contradict this, however: in a spontaneous speech material of nearly two hundred speakers, the speech tempo of forty-year-olds turned out to be the fastest, as opposed to both younger and older groups of speakers (Jacewicz et al., 2010). Significant differences were found between the speech rates of neutral spoken texts vs. ones produced in various joyful or sorrowful states of mind (Schnoebelen, 2010). An increase of the speech rate may be caused by the fact that the speaker considers the given portion of the message less important; but it can also be due to some external factor like the behavior of the interlocutor. The transformation of the speaker's ideas into speech may become slower due to conceptual planning becoming hesitant, construction of the utterance becoming difficult, or lexical selection becoming riddled by competitive lexemes at the given point. In the phrases of spontaneous Italian narratives, the tempo of syllables has been measured, and compared between pre-stress and post-stress positions (Cutugno and Savy, 1999). The results showed that after phrasal stress, the tempo increased (by some 65%), while in pre-stress positions, such increase was only by 33%. The decrease of speech rate, on the other hand, where it occurred, was 15% in a post-stress position and 40% before the stressed syllable. It can be concluded that the temporal properties of a longer stretch of spontaneous speech are not constant and not independent of other prosodic properties of speech like stress, or intonation (Keller and Port, 2007).

Inter-speaker variation is significant; but large variability can also be found across utterances of one and the same speaker. In spontaneous English conversations, for instance, 33% large changes were attested in speech rate with one of the speakers (Chafe, 2002).

Data from perceptual experiments make it probable that speakers tend to employ general features as boundary markers of thematic units (TU) and of phrases, ones that can also be used in decoding. Thematic units are portions of discourse exhibiting coherence of content that are appropriately structured both syntactically and prosodically (Swerts et al., 1992; Georgakopolou and Goutsos, 2004). In determining phrases within spontaneous narratives or dialogues, on the other hand, primarily rises and falls of speech melody, as well as stress relationships are taken into consideration (Botinis

et al., 2003). So-called idea units (brief coherent spontaneous text segments) are taken to be 2 seconds long on average, corresponding to roughly 6 English words.

The findings of the present research will throw new light on temporal properties of spontaneous narratives, on covert processes of speech planning and pinpoint universal and individual characteristics, features characterizing several speakers and single speakers, respectively. We hypothesize that (i) spontaneous narratives can be segmented into units defined by acoustic-phonetic parameters: these are thematic units that are further segmentable into phrases, (ii) phrases exhibit characteristic temporal patterns, and (iii) thematic units are mostly universal but can also be taken to be based on individual peculiarities to some extent.

2. Subjects, material, method

For this study, we used 10 interviews of the BEA Hungarian Spontaneous Speech Database (Gósy, 2012) in which the participants talk about their job, family, and hobbies. Five of the speakers are female, and five are male; all of them native speakers of Hungarian from Budapest; aged between 22 and 35.

The total material is 57 minutes long (3–8 minutes per informants), and was annotated in Praat 5.1 (Boersma and Weenink, 2010) at several levels (thematic units and phrases encoded orthographically and in phonetic transcription, and sound-level annotation). In the case of voiced segments, the first period was taken to be the boundary. Using a Praat script, we automatically extracted fundamental frequency (F0) and intensity. (We sampled both at every 200 ms.) The initial criterion of the definition of thematic units (TU) was that the interviewer opened a new topic by each question, that is, the preceding portion of text was a unit semantically, syntactically, and prosodically, as well. The interviewer started a new topic only when the speaker indicated, verbally or in some other manner, that s/he did not want (or could not) say anything more. Within thematic units, we separated phrases by either or both of the following two criteria: (i) an utterance flanked by (silent or filled) pauses on both sides, and/or (ii) a radical change in fundamental frequency and intensity.

We automatically determined the occurrence and duration of all labeled silent and filled pauses, and of all phrases, and calculated automatically the rate of articulation, defined as the number of segments per total articulation time. The corpus included a total of 7863 words. The informants uttered an average of 177 words per minute. For statistical analyses, we used the SPSS 13.0 program (analysis of variance, correlation analysis).

2. Results

Description of the results will be organized in four subsections of temporal analysis which concern silent and filled pauses, temporal properties of thematic units, and phrases as well as articulation tempo.

2.1. Silent and filled pauses

Our analyses have confirmed that phrases can be reliably defined in terms of pauses. The corpus included 1326 silent pauses, of a mean duration of 510 ms (SD: 405 ms). The shortest pause took 23 ms, and the longest took 3036 ms. The number of pauses found with individual speakers exhibited extensive variability (Fig. 1). The duration of silent pauses was significantly different across speakers ($F(9,1326) = 17.422$; $p < 0.001$) but a post-hoc test showed that the difference was only significant between two speakers and all the others.

The number of filled pauses was 260 in the corpus. Their mean duration was 323 ms (SD: 153 ms). The shortest filled pause took 20 ms, and the longest one took 720 ms. One-way ANOVA confirmed significant differences across speakers ($F(9,219) = 6.704$; $p < 0.001$), but a post-hoc test showed that the difference was only significant between a single speaker (speaker 4 in Fig. 1) and all the others. Correlation analysis showed that pausing exhibited individual differences across speakers; if the speech of a speaker was characterized by longer silent pauses, s/he also tended to produce longer filled pauses ($R^2 = 0.643$; $p = 0.045$).

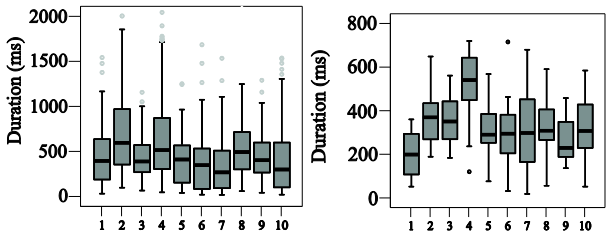

Fig. 1: Duration of silent (left panel) and filled pauses (right panel) (1-5=females, 6-10=males)

2.2. Temporal properties of thematic units

With 60% of the speakers, the narrative could be segmented into three thematic units; the rest of the speakers produced 5 or 6 thematic units. Starting a new topic as the criterion for thematic unit boundaries was correlated with changes in fundamental frequency and intensity; thus, TU boundaries were predictable.

The mean duration of TUs was 56 s (SD: 48 s). The distribution of durations was lognormal (Fig. 2), meaning that most duration figures fell between zero and 100 s, and that the curve decreased in a protracted manner.

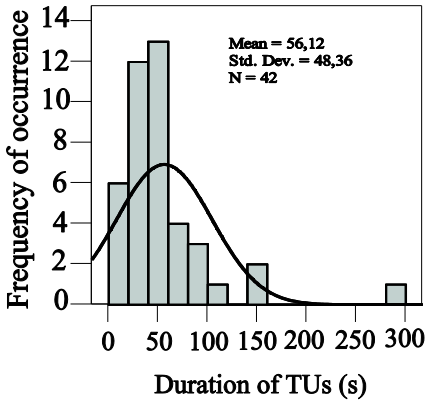

Fig. 2: The distribution of duration of TUs

In the duration of thematic units, with two exceptions, there were no significant differences across speakers (Fig. 3). TU durations of speakers 2 and 3 significantly differed, according to post-hoc tests, from the data of all the other speakers (one-way ANOVA: $F(9,302) = 5.485$;

$p < 0.001$). These informants produced far longer thematic units than the others did.

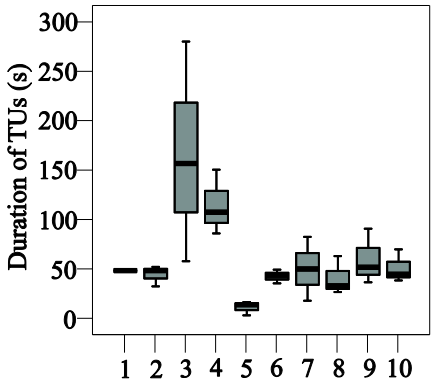

Fig. 3: The duration of TUs in individual speakers' narratives (1-5=females, 6-10=males)

The position of TUs within the narratives may have influenced their duration. For an analysis of this, we only considered narratives that contained three thematic units, given that the duration of these units did not exhibit significant differences. The trend was that TUs get shorter as the end of the narrative draws nearer (Fig. 4).

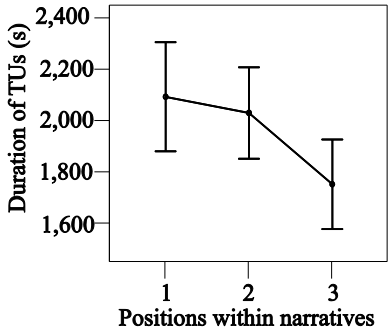

Fig. 4: Duration of TUs in various positions within narratives (1 = initial; 2 = medial; 3 = final)

Hungarian speakers produce almost 20 words less in a minute than English speakers do; the relevant figure for English is 196 words per minute (Youan et al., 2006). This difference is obviously due to the fact that Hungarian, being an agglutinative language, has longer words (the average syllable count of Hungarian words in spontaneous speech is 3.5). The mean number of words per thematic unit was 245 (SD: 199), irrespective of whether they were content words or function words.

Statistical analysis showed no differences in word count depending on which part of the narrative contained the given TU. The smallest number produced in a thematic unit was 147 words per minute, and the largest was 206 words per minute.

2.3. Temporal properties of phrases

The number of phrases was 1394 in our material. Their number within TUs was not independent of whether the TU was initial, medial, or final in the narrative. Medial thematic units consisted of fewer phrases than the preceding or following ones (Fig. 5). The duration differences of phrases within thematic units were significant (one-way ANOVA $F(9,1394) = 11.175$; $p < 0.001$). Their variability was larger across speakers than that of the duration of thematic units.

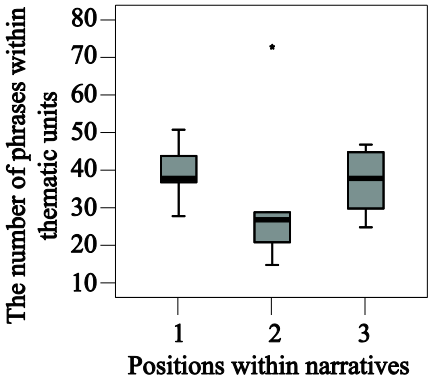

Fig. 5: The number of phrases within thematic units (in six speakers' material)

Speakers can be classified into two groups, one group produced relatively short phrases, while the other group produced relatively long ones. The position of thematic units within narratives also affected the length of phrases (Fig. 6).

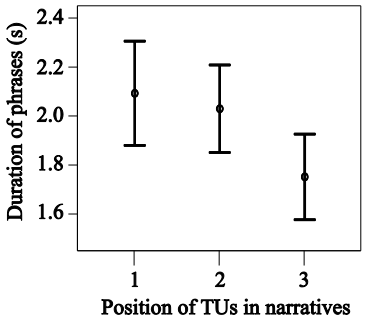

Fig. 6: The duration of phrases in terms of the position of TUs (1= initial; 2=medial; 3=final)

Narrative-final TUs were realized in shorter duration than the preceding ones (one-way ANOVA $F(2,750) = 3.277$; $p = 0.038$). The average word count in phrases within thematic units was 5.8 words (SD: 4.7, minimum: 3.4, maximum: 8.1).

The average word count of phrases is lognormal, and exhibited significant differences depending on which TU the given phrase occurred in. The phrases of third thematic units contained fewer words on average than those of first and second ones (1st TU = 6.2 words; 2nd TU = 6.1 words; 3rd TU = 5.1 words; one-way ANOVA: $F(2,750) = 4.313$; $p = 0.014$). That is, towards the end of a narrative, it was not only the case that the thematic units got shorter, but also the phrases they contained were shorter and consisted of fewer words. We found strong linear correlation between the number of words in a phrase and its duration ($R^2 = 0.8603$; $p < 0.001$). This means that the longer the duration of a phrase the more words it consists of (Fig. 7).

2.4. Rate of articulation

The slowest speaker exhibited a mean rate of articulation of 11.7 sounds/s (SD: 3.1), the fastest speaker exhibited 15.4 sounds/s (SD: 6.5). Statistical analyses confirmed significant differences of rate of articulation across

speakers (one-way ANOVA: $F(9,1387) = 13.168$; $p < 0.001$). However, a post-hoc test showed that three speakers differed from nearly all other speakers.

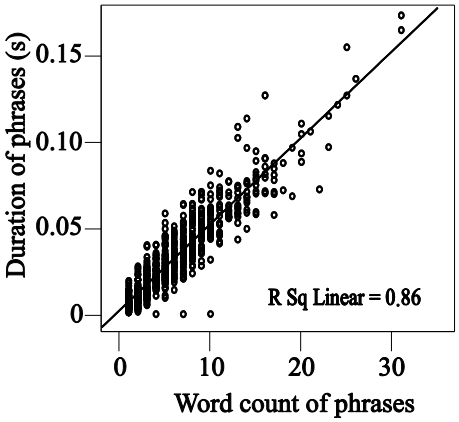

Fig. 7: The correlation between the duration and word count of phrases

Among speakers producing three thematic units, we found two different tendencies in tempo changes across TUs. With three of them, the mean rate of articulation accelerated in the second TU compared to the first, and then got slower toward the end of the narrative. With the other three, on the contrary, the rate of articulation was slower in the second TU than in the first, and then a strong acceleration occurred toward the end of the narrative (Fig. 8).

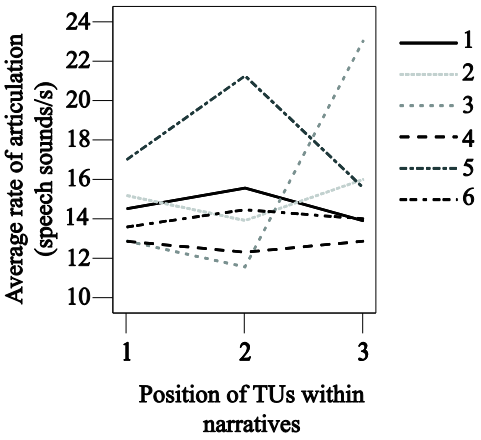

Fig. 8: Average rate of articulation in individual TUs

Given that the rate of articulation changes continuously in the narratives, we performed continuous time analysis of the rate of articulation of phrases. As compared to the mean rate of articulation of the whole narrative, extremely fast and extremely slow values were both found in the individual phrases (Fig. 9).

3. Conclusions

Spontaneous speech corpora make it possible to perform a thorough analysis of temporal properties of spontaneous speech. The mean tempo values can only be a point of departure, followed by detailed analyses of the complex temporal patterns of spontaneous utterances. In the present series of investigations, we determined thematic units and phrases, and gave objective values of the parameters measured. We found that (i) the majority of speakers (60% in our case) organized their narratives in similar temporal structures, (ii) thematic units could be identified in terms of certain prosodic criteria, (iii) we found statistically valid correlations across factors like the duration of phrases, the word count of phrases, the rate of articulation of phrases, and pausing characteristics, and (iv) these parameters exhibited extensive variability both across and within speakers.

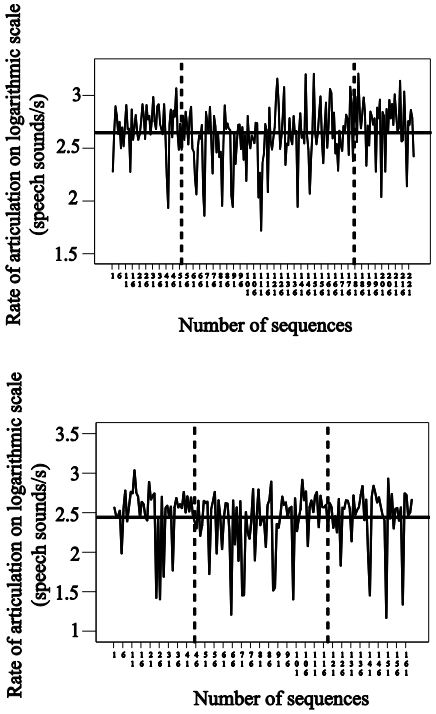

Fig. 9: Rate of articulation in two speakers' narratives (the horizontal line represents the average rate of articulation of the whole narrative; the vertical lines indicate the boundaries of TUs)

According to our data, speakers create thematic units of roughly similar duration in their narratives, that is, we can assume the existence of a kind of "internal time control" as part of covert speech planning processes that determines how long speakers may dwell on a given topic in a non-conversational situation. This control function probably takes several factors into consideration, including the listener's assumed level of interest, the amount of information to be shared with the interlocutor, selection, avoidance of certain details, etc. While filled pauses did not differ in length in a statistically relevant manner, silent pauses did. This can be due to physiological factors like the regulation of breathing, but obviously a number of other factors play a role in how long silent pauses a speaker produces. Pauses, being generally accepted boundary markers, appear to be language specific in both their occurrence and phonetic properties (Zellner, 1994; Tseng, 2006). Narrative-medial TUs tend to consist of fewer phrases than the TUs before and after them. This can be due to the fact that the speaker tends to elaborate the first topic in relatively more detail, requiring more thought and speech planning, a fact that emerges in the production of a higher number of phrases. In the second topic, the speaker employs strategies of narrative construction more easily, speaks

more concisely, and produces fewer phrases. In the case of the third topic, however, the speaker appears to lose interest, find solitary speech production inconvenient, or simply get tired, given that in everyday communication the construction of lengthier narratives is not typical.

All those factors may result in the fewer phrases that characterize the last TUs of narratives. The objective temporal data reflect the same pattern. Rate of articulation is expected to exhibit great variability both across and within speakers. The rate of articulation of individual speakers follows two clear tendencies, in which the second thematic unit has a crucial role. But the appearance of extreme values characterizes all phrases.

Our first hypothesis, according to which units defined by acoustic-phonetic parameters can be determined within spontaneous narratives, was confirmed. Thematic units were getting shorter towards the end of the narratives, whereas in terms of the number of words involved, there was no statistically confirmed difference across TUs.

Our second hypothesis was that the phrases making up the thematic units would exhibit particular temporal patterns. This was also confirmed. The duration of phrases showed a lot more variability across speakers than that of thematic units did. It appears, then, that phrases primarily exhibit speaker-dependent properties. Their duration is affected by where exactly they occur within a thematic unit. A strong correlation was found between the number of words in a phrase and its duration, confirming the claim that in longer phrases the speaker indeed produces more words than in shorter ones.

In our third hypothesis, we stated that the properties of thematic units are universal to a larger extent than they are speaker specific. On the basis of our results, this statement has to be qualified. Although the temporal organization of narratives exhibits a number of universal properties, individual properties may override these in interesting ways (Russo and Barry, 2008).

Narrative-internal tempo changes may depend on a number of further factors. The present paper demonstrated some objective characteristics of the ways narratives are organized, including properties that are true of speakers in general and those that characterize them individually.

References

Boersma, P. and Weenink, D. (2010). *Praat: doing phonetics by computer*. (http://www.fon.hum.uva.nl/praat/download_win.html, Access date: January 15, 2011.)

Botinis, A., Gawronska, B., Katsika, A. and Panagopoulou, D. (2003). Prosodic speech production and thematic segmentation. In: *PHONUM*, 9, pp. 113-116.

Chafe, W. (2002). Prosody and emotion in a sample of real speech. In: Fries, P. H., Cummings, M., Lockwood, D., Spruiell, D. (Eds.) *Relations and functions within and around language*. London: Continuum. pp. 277-315.

Cutugno, F. and Savy, R. (1999). Correlation between segmental reduction and prosodic features in spontaneous speech: the role of tempo. In: *Proceedings of the XIVth International Conference of the Phonetic Sciences*. San Francisco. pp. 471-474.

Georgakopolou, A. and Goutsos, D. (2004). *Discourse analysis: an introduction*. Edinburgh: Edinburgh University Press.

Gósy, M. (2012). BEA - a multifunctional Hungarian spoken language database. In: *The Phonetician*, pp. 51-62.

Jacewicz, E., Fox, Allen, R. and Lai, W. (2010). Between-speaker and within-speaker variation in speech tempo of American English. In: *Journal of the Acoustical Society of America*, 128, pp. 839-850.

Keller, E. and Port, R. (2007). Speech timing: Approaches to speech rhythm. In: *XVIth International Conference of the Phonetic Sciences*. Saarbrücken. pp. 327-329.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. In: *Journal of the Acoustical Society of America*, 59, pp. 1208-1221.

Quené, H. (2005). Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In: *Proc. of Interspeech 2005*. Lisbon, Portugal. pp. 2457-2460.

Russo, M. and Barry, W. J. (2008). Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo. In: *Proc. of Speech prosody 2008*. (http://www.iscaspeech.org/archive. Access date: December 4, 2010)

Schnoebelen, T. (2010). Variation in speech tempo: Capt. Kirk, Mr. Spock, and all of us in between. In: *Proc. of 36th Conference on New Ways of Analyzing Variation: Diversity, Interdisciplinarity, Intersectionality*. San Antonio, Texas, December 2010.

Swerts, M., Geluykens, R. and Terken, J. (1992). Prosodic correlates of discourse units in spontaneous speech. In: *Proceedings of the International Conference on Spoken Language Processing*. Banff. pp. 421-424.

Tseng, S-Ch. (2006). Linguistic markings of units in spontaneous Mandarin. In: Huo, Q., Ma, B., Chang, E-S., Li, H. (Eds.) *Chinese spoken language processes. Proceedings of the 5th International Symposium*. Singapore: Springer. pp. 43-54.

Youan, J., Liberman, M. and Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. Proceedings of the 9th International Conference on Spoken Language Processing. Pittsburgh, PA, USA. (http://www.isca-speech.org/archive. Access data: December 7, 2010)

Verhoeven, J., De Pauw, G. and Kloots, H. (2004). Speech rate in a pluricentric language: a comparison between Dutch in Belgium and the Netherlands. In: *Language and Speech*, 47, pp. 297-308.

Zellner, B. (1994). Pauses and the temporal structure of speech. In: Keller, E. (Ed.) *Fundamentals of speech synthesis and speech recognition*. Chichester: John Wiley. 41-62.

This research was supported by OTKA No. 108762.